Affect-aware Cross-Domain Recommendation for Art Therapy via Music Preference Elicitation

Bereket A. Yilma
Luis A. Leiva
name.surname@uni.lu
University of Luxembourg
Luxembourg

Abstract

Art Therapy (AT) is an established practice that facilitates emotional processing and recovery through creative expression. Recently, Visual Art Recommender Systems (VA RecSys) have emerged to support AT, demonstrating their potential by personalizing therapeutic artwork recommendations. Nonetheless, current VA RecSys rely on visual stimuli for user modeling, limiting their ability to capture the full spectrum of emotional responses during preference elicitation. Previous studies have shown that music stimuli elicit unique affective reflections, presenting an opportunity for cross-domain recommendation (CDR) to enhance personalization in AT. Since CDR has not yet been explored in this context, we propose a family of CDR methods for AT based on music-driven preference elicitation. A large-scale study with 200 users demonstrates the efficacy of music-driven preference elicitation, outperforming the classic visual-only elicitation approach. Our source code, data, and models are available at https://github.com/ArtAICare/Affect-aware-CDR.

CCS Concepts

• **Information systems** → **Personalization**; **Recommender systems**; • **Computing methodologies** → *Learning latent representations*; • **Applied computing** → *Media arts*.

Keywords

Recommendation; Personalization; Artwork; User Experience; Machine Learning

1 Introduction

Art therapy (AT) has long been recognized as a powerful intervention for psychological distress, including trauma, anxiety, and depression [32]. It provides individuals with a medium for creative expression that facilitates emotional processing and self-exploration, fostering recovery and well-being. Particularly, a visual exposure

form of AT includes a preference elicitation phase, where patients select paintings from a curated set of visual stimuli that resonate with their healing journey, followed by a guided session featuring a similar set of artworks recommended by art therapists, designed to evoke positive emotions and contain healing elements [17].

A significant challenge in AT lies in personalizing these interventions for individuals with unique traumatic experiences. Traditional AT requires the expertise of trained therapists in selecting artworks that are therapeutically appropriate for each patient's healing journey. This process, while effective, is resource-intensive and constrained by the therapist's familiarity with available artworks and the diversity and uniqueness of patients' emotional needs. The vastness of the search space for relevant visual stimuli makes it difficult to identify the right pieces for each patient, posing a substantial barrier to scalable, personalized interventions [18].

Recent advancements in digital AT have aimed to improve both the accessibility and the effectiveness of therapeutic interventions through computational tools. In particular, Visual Art Recommender Systems (VA RecSys) have recently been developed to personalize visual stimuli based on the psychological states of individuals, showing promising results. For example, Yilma et al. [49, 50] developed VA RecSys that promote healing and recovery in the context of Post-Intensive Care Syndrome (PICS) intervention. While useful, they only considered visual stimuli for preference elicitation. However, a different modality, music, has the potential to further elicit affective cues that can be missed in visual stimuli [14, 24]. Music plays a significant role in eliciting emotional responses [39, 45], which are essential for psychological healing, potentially enhancing the personalization and effectiveness of digital AT.

State-of-the-art (SOTA) approaches in digital AT have not yet explored the cross-domain recommendation (CDR) potential of affective stimuli knowledge transfer between VA and music. Hence, we propose a novel approach that integrates music as a preference elicitation modality for personalized VA RecSys in AT. Concretely, we propose a family of CDR methods: (1) *Mozart*, or *affect-aware contrastive alignment of music and paintings*, which maps the affective properties of music and paintings into a shared latent space that allows a semantic retrieval of therapeutic content across the two modalities; (2) *Haydn*, or *affective space search*, based on k-nearest neighbors, which leverages the valence and arousal labels of each modality to construct a similarity matrix for retrieval; (3) *Salieri*, or *multimodal alignment of music and paintings*, using large language models (LLMs) and vision-language models (VLMs), which enables the extraction of semantic representations from both modalities, facilitating cross-modal mapping for retrieval. Our main research

question is: *Can music-driven preferences provide therapeutic art recommendations?*

We assess the efficacy of these CDR strategies for AT on a large-scale user study against the uni-modal SOTA VA RecSys engine reported in previous work, comprising a ResNet-50 backbone and visual-only preference elicitation. In sum, we make the following contributions:

- We introduce CDR to the field of AT by integrating music and VA domains.
- We develop and evaluate three novel affect-aware CDR algorithms based on music preferences.
- We conduct a large-scale user study with 200 people with psychiatric sequelae, comparing the proposed approaches against the SOTA baseline.
- We provide insights on integrating CDR into digital AT and discuss future research for mental health interventions.

2 Background and Related Work

2.1 VA RecSys in Art Therapy

VA pieces like paintings have been widely utilized to promote psychological well-being in clinical environments [21]. As trauma can often be hard to verbalize in spoken psychotherapy sessions, art opens a creative medium of expression [12], stimulating emotions and self-reflection, and addressing various psychological disorders such as depression and anxiety [17, 43].

To offer an effective AT based on the unique needs of patients, personalization remains vital yet largely constrained due to its reliance on human expertise and a manual art selection process [1]. Very recently, VA RecSys have emerged to support the delivery of personalized AT. Yilma et al. [49] developed VA RecSys that leverage visual as well as textual features of paintings for semantic retrieval of therapeutic art recommendations. The proposed models used Residual Neural Network (ResNet50) [19] for image feature extraction, Latent Dirichlet Allocation (LDA) [4] and Bidirectional Encoder Representations from Transformers (BERT) [11] for textual feature extraction, and Bootstrapping Language-Image Pretraining (BLIP) [28] for multimodal feature extraction. A follow-up study [50] investigated a human-in-the-loop approach to safeguard recommendations and foster human-AI collaboration in AT. While these systems have shown promise, they rely solely on visual inputs, potentially missing broader affective dimensions. On the other hand, music serves as a complementary modality for preference elicitation, owing to its capacity to elicit emotional responses [24]. Several studies have also linked music to therapeutic neurological responses, suggesting its role in preference detection.[6, 25, 46].

Jazi et al. [51] proposed an affect-aware music RecSys that maps user interactions to emotional states, while Patel et al. [36] and Iordanis [22] developed Deep Learning models for personalized music recommendations based on affective cues. These works highlight music's potential to reveal affective preferences that could inform visual art recommendations. However, music as a preference elicitation tool for VA RecSys remains unexplored, motivating our investigation into a CDR approach in AT via affective knowledge transfer between music and paintings.

2.2 CDR and the Music-Art Gap

Cross-domain recommendation leverages user preferences or item features from one domain to enhance recommendations in another domain, mitigating issues like data sparsity and cold-start problems [13, 52]. CDR has been particularly useful in the context of movie recommendations, using for example book [54] and music [15] preferences. The key idea is to find a suitable embedding and mapping framework [30, 33, 53]. Recent works have also explored graph-based methods for CDR [20]

Despite these advances, CDR between music and VA remains underexplored, and in particular in therapeutic contexts. A study by Campos-Bueno et al. [7] investigated the affective dimensions of valence and arousal in music and VA. They found that music elicited stronger effects on valence than paintings. It has been reported that both modalities elicit profound emotional responses, music through auditory stimulation [45] and VA via imagery [17], yet no framework transfers affective knowledge between them for CDR. Lee et al. [27] investigated music-to-art style transfer using generative adversarial networks, but their focus was aesthetic rather than therapeutic.

Emotional responses from music have long been leveraged in music therapy for emotional regulation, where specific musical elements such as tempo, melody, and harmony are known to influence affective states [24, 39]. Recent advancements have applied machine learning techniques to predict the effectiveness of music therapy interventions [23, 38]. Similarly, paintings have been shown to evoke strong emotional reactions, with different artistic styles, color palettes, and subject matter being used in AT to help individuals process trauma, anxiety, and depression [17, 32]. Further research in neuroscience underscores the role of affective control in emotion regulation, which can support therapeutic outcomes [26, 42]. Recent studies found that active and passive engagement in creative arts activates neural circuits that promote adaptive emotional regulation [3]. Scoping reviews of integrated arts therapies and neuroscience emphasize the need for interdisciplinary approaches to compare modalities like music and visual arts [5].

In sum, it is evident that music has potential to evoke profound affective responses. Hence, our main hypothesis is that the integration of affective cues across these two domains can derive better therapeutic content recommendations. Unfortunately, no affect-aware CDR framework has been proposed to date in a unified therapeutic context. To address this critical gap, we propose a family of CDR approaches that leverage music clips for preference elicitation to derive therapeutic art recommendations.

3 Method: Therapeutic VA RecSys from Music

In this work, we depart from utilizing visual stimuli for preference elicitation. Instead, we leverage music stimuli to elicit affective preferences for therapeutic painting recommendations. To the best of our knowledge, no prior work in digital AT involves cross-domain affective knowledge transfer. Thus, we take the first step to introduce affect-aware CDR in digital AT.
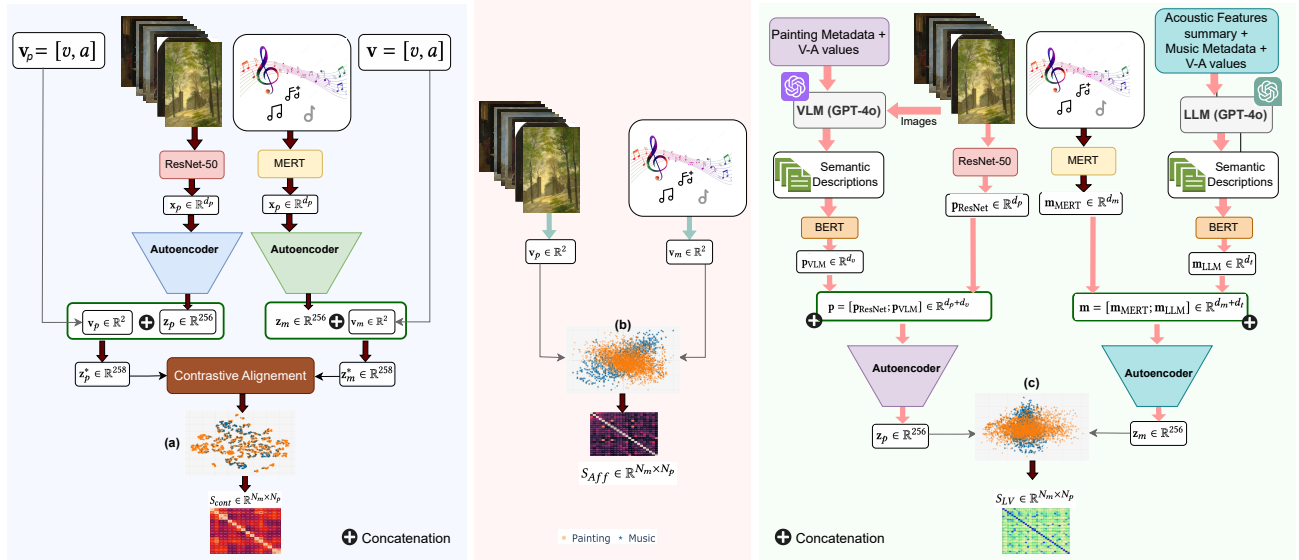
Figure 1: Proposed system architectures. From left to right: Affect-aware contrastive alignment (Mozart), Affective Space Search (Haydn), and multimodal alignment with LLM and VLM (Salieri). Notice that each approach produces different latent embedding spaces (a,b,c).

3.1 Dataset

Although evidence from both music and AT literature highlights the potential for affect-aware CDR in AT, a significant challenge remains in the availability of parallel music and VA datasets with affective annotations. These datasets need to be carefully curated to ensure they match therapeutic goals and reflect the emotional qualities of each modality. While there are several datasets for music and paintings individually, few are integrated with affective labels, making it difficult to leverage them for the affect-aware CDR task. To overcome this limitation, we conducted a thorough analysis and selected reliable public datasets from both domains that are well suited for this task. In the following, we outline the dataset selection, preprocessing steps, and the proposed CDR strategies.

3.1.1 Music Dataset. We selected the DEAM (Database for Emotional Analysis of Music) dataset,[1] as it has been reliably used in music therapy [8, 35]. DEAM contains 1,802 instrumental music excerpts annotated with valence and arousal values both continuously (per-second) and over the whole song. For our purposes, we used the averaged per-song annotations but filtered out songs with high standard deviation to enforce stability and reduce ambiguity. We excluded songs with valence SD > 1.75 and arousal SD > 1.0, chosen based on the label distributions. This filtering resulted in 909 songs.

3.1.2 Visual Art Dataset. For paintings, we chose the WikiArt Emotions Dataset,[2] as it is the only public dataset with reliable affective annotations. It contains 4,105 paintings annotated based on combined image and metadata analysis. Each painting is labeled with multiple emotions and their intensities.

3.1.3 Preprocessing. Both datasets have different labels to represent their emotional triggers on users. DEAM is annotated with continuous valence-arousal (V-A) scores in the $[-1, 1]$ range, while WikiArt is labeled with multiple emotions and their intensities. First, we need to convert the labels of the two datasets to uniform affective annotations (i.e., V-A values). For DEAM, we considered per-song static annotations and filtered out songs with high standard deviation, to ensure stability and reduce ambiguity. Then, to align the WikiArt dataset emotional labels with the V-A annotations, we first normalized the intensity of each emotion label for each painting and then we used the 'NRC Valence, Arousal, and Dominance' (NRC VAD) Lexicon, which provides V-A scores for over 20,000 English words,[3] to translate these emotions into V-A scores, in the $[-1, 1]$ range.

After this, we obtained two modalities, paintings and music; both with respective modality-specific features and V-A values. The final dataset is comprised of 909 music samples and 4,105 painting samples.

Although these affective labels are considered to be representative, we hypothesize that the feature-level nuances of each modality can play a significant role in the emotional reflection triggered by each modality [2, 9]. Therefore, we aim to learn a joint representation space guided by emotional awareness to derive better recommendations.

3.2 Mozart: Cross-Domain Recommendation via Affect-Aware Contrastive Alignment

To establish a unified representation space, we start by extracting modality-specific features using pre-trained networks: MERT [29]

---

for music and ResNet50 [19] for paintings. Let $\mathbf{x}_m \in \mathbb{R}^{d_m}$ and $\mathbf{x}_p \in \mathbb{R}^{d_p}$ be the extracted feature vectors for music and paintings, respectively, where d_m and d_p denote the dimensionality of the raw feature representations from each modality. We then obtain reduced representations through modality-specific autoencoders:

$$\mathbf{z}_m = AE_m(\mathbf{x}_m), \quad \mathbf{z}_p = AE_p(\mathbf{x}_p), \tag{1}$$

where AE_m and AE_p are autoencoders trained with reconstruction loss separately for music and paintings, respectively, to learn compact 256D embeddings: $\mathbb{R}^{256}$. This ensures that the embeddings retain meaningful information while reducing dimensionality.[4] Thus, producing embeddings $\mathbf{z}_m, \mathbf{z}_p \in \mathbb{R}^{256}$.

Since emotional perception is fundamental to both music and VA, we explicitly incorporate V-A representations. Each sample is associated with a V-A vector $\mathbf{v}_i = [v_i, a_i]$, where v_i and a_i denote the valence and arousal values, respectively. Finally we concatenate the raw V-A values with the learned embeddings:

$$\mathbf{z}_m^* = [\mathbf{z}_m, \mathbf{v}_a], \quad \mathbf{z}_p^* = [\mathbf{z}_p, \mathbf{v}_a]. \tag{2}$$

This results in enriched embeddings $\mathbf{z}_m^*, \mathbf{z}_p^* \in \mathbb{R}^{258}$ that encode both modality-specific features and affective information, forming the basis for the subsequent contrastive learning objective.

3.2.1 Contrastive Learning with Continuous V-A Values.
Traditional contrastive learning frameworks rely on binary labels to indicate whether sample pairs are similar or dissimilar. In our work, each sample of music and painting is associated with continuous V-A values. To effectively capture the degree of emotional similarity, we first compute the Euclidean distance[5] between the V-A vectors of two samples:

$$d_{ij} = \|\mathbf{v}_i - \mathbf{v}_j\|_2 = \sqrt{(v_i - v_j)^2 + (a_i - a_j)^2}, \tag{3}$$

where $\mathbf{v}_i = [v_i, a_i]$ represents the V-A vector for sample i. To convert this distance into a soft, continuous similarity measure, we apply a Gaussian kernel:

$$S_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right), \tag{4}$$

where σ is a bandwidth parameter that controls how quickly similarity decays with increasing distance. This approach is particularly effective for continuous labels, as it allows the model to weigh the contribution of each pair according to the degree of emotional closeness. Recent work by Tan et al. [44] demonstrates that kernel-based similarity measures enhances the learning of continuous representations in cross-modal settings. In our work, the continuous similarity score S_{ij} is integrated into the contrastive loss:

$$\mathcal{L} = \sum_{i,j} S_{ij} \left(\|\mathbf{z}_i^* - \mathbf{z}_j^*\|_2\right)^2 + (1 - S_{ij}) \max\left(0, m - \|\mathbf{z}_i^* - \mathbf{z}_j^*\|_2\right)^2, \tag{5}$$

where $\mathbf{z}_i^*$ and $\mathbf{z}_j^*$ are the enriched joint embeddings incorporating both modality-specific features and V-A information. The margin m ensures a minimum separation between dissimilar samples.

---

[4]256D was chosen to balance expressiveness and efficiency. Higher dimensions increase model complexity and computational cost, while lower dimensions may not sufficiently capture feature-rich representations. An offline evaluation (ranking overlap) showed no significant gains with 512D or 1024D.
[5]Unlike cosine similarity, which ignores magnitude, Euclidean distance ensures that emotionally distinct but directionally similar items are properly differentiated.

3.2.2 Addressing Dataset Imbalance.
The disparity in dataset sizes 909 music samples versus 4,105 art samples can bias the learning process. To mitigate this imbalance, we adopted a weighted loss approach. We adjust (5) by assigning weights to each sample based on its modality. Define the weighting factors as:

$$\lambda_m = \frac{N_p}{N_m + N_p}, \quad \lambda_p = \frac{N_m}{N_m + N_p}, \tag{6}$$

where $N_m = 909$ and $N_p = 4105$. The weighted contrastive loss then becomes $\mathcal{L}_{\text{weighted}} = \sum_{i,j} \lambda_i \lambda_j \mathcal{L}$,

where λ_i equals λ_m if sample i is from the music domain and λ_p if from the art domain.

3.2.3 Implementation details.
Music features were extracted using MERT-v1-330M.[6] Audio files were resampled to 24 kHz and converted to mono. The MERT-based embeddings were concatenated with precomputed acoustic features (means, medians, and standard deviations of tempo, spectral contrast, etc.) from DEAM. For paintings, a pre-trained ResNet50 model extracted 2048D features from images resized to 224×224 pixels, using ImageNet-standard normalization, sourced from the WikiArt dataset. Music and painting embeddings were further reduced to 256D using modality-specific autoencoders with encoder layers `[input_dim, 1024, 512, 256]` and mirrored decoders, trained with Adam optimizer ($\eta = 10^{-3}, \beta_1 = 0.9, \beta_2 = 0.999$) and MSE loss. The resulting 256D embeddings were Min-Max scaled to $[-1, 1]$ and concatenated with the V-A labels from each dataset, producing 258D enriched embeddings $(\mathbf{z}_m^*, \mathbf{z}_p^*)$ per modality. Then, for contrastive alignment, a projection head, implemented as a two-layer MLP with architecture $[258, 256, 128]$, mapped the 258D embeddings to a 128D joint space.

The modality-specific weights of the weighted contrastive loss are $\lambda_m = 0.818$ and $\lambda_p = 0.182$, based on the weighting factors in Section 3.2.2. The optimal hyperparameters for the Gaussian kernel-based similarity measure (Equation 4), i.e., the bandwidth parameter σ and the margin m, were determined via a grid search. The best performance was achieved with $\sigma = 0.5$ and $m = 0.5$. Mozart was trained for up to 50 epochs using Adam optimizer ($\eta = 10^{-3}, \beta_1 = 0.9, \beta_2 = 0.999$) and early stopping of 5 epochs, monitoring the validation loss. The resulting joint embedding space, visualized in Figure 1, Before, the similarity matrix shows a rather effective alignment, where both modalities contribute to shaping the learned representation guided by a combination of affective and modality-specific features.

By computing Euclidean distances between the 128D joint embeddings, we get a similarity matrix $\mathbb{R}^{N_m \times N_p}$ between all music and painting pairs, enabling efficient retrieval of painting recommendations for music tracks. Then, given a user's list of preferred music tracks $m_1, m_2, \ldots, m_k$ with corresponding ratings $r_1, r_2, \ldots, r_k$ on a 5-point Likert scale, we normalize the ratings to obtain weights $w_1, w_2, \ldots, w_k$, where each $w_i \in [0, 1]$ and $\sum_{i=1}^k w_i = 1$. The weighted average distance from the set of music tracks to each painting p_j is then calculated as:

$$d_{\text{Mozart}}(p_j) = \sum_{i=1}^k w_i \cdot \|\mathbf{z}_{m_i}^+ - \mathbf{z}_{p_j}^+\|_2 \tag{7}$$

---

[6]https://huggingface.co/m-a-p/MERT-v1-330M

where $\mathbf{z}_{m_i}^+$ and $\mathbf{z}_{p_j}^+$ are the 128D joint embeddings. The top-n paintings with the smallest $d_{\text{Mozart}}(p_j)$ values are recommended, reflecting their emotional similarity to the user's music preferences, weighted by the normalized ratings.

3.3 Haydn: Cross-Domain Recommendation via Affective Space Search

We implemented a k-nearest neighbors (kNN) VA RecSys approach using affective space search which allows retrieval based on the affective alignment of music and painting V-A vectors. Each music track m_i and painting p_j is represented by its V-A vectors $\mathbf{v}_{m_i}$ and $\mathbf{v}_{p_j}$, respectively. The affective distance between them is computed using the Euclidean distance, as defined in Equation 3, yielding a similarity matrix $\mathbb{R}^{N_m \times N_p}$ between all music and painting pairs.

Similar to Mozart, a weighted average is computed to make recommendations from a precomputed similarity matrix derived from Euclidean distances, with weights $w_i \in [0, 1]$ and $\sum_{i=1}^{k} w_i = 1$. The weighted average affective distance from the set of music tracks to each painting p_j is calculated as:

$$d_{\text{Haydn}}(p_j) = \sum_{i=1}^{k} w_i \cdot \|\mathbf{v}_{m_i} - \mathbf{v}_{p_j}\|_2 \tag{8}$$

where $\mathbf{v}_{m_i}$ and $\mathbf{v}_{p_j}$ are the V-A vectors of music m_i and painting p_j respectively. This is a straightforward and interpretable method, though it may not fully capture modality-specific attributes responsible for the different emotional reflections triggered on users.

3.4 Salieri: Cross-Domain Recommendation via multimodal Alignment with LLM and VLM

We propose a third approach based on large language models (LLMs) and vision-language models (VLMs), as these models demonstrate a strong capability in capturing subtle emotional and semantic details, which are crucial for affective cross-modal retrieval. Salieri provides a complementary perspective to both the affective space search and contrastive alignment approaches discussed earlier. This approach constructs multimodal representations for both music and paintings by leveraging transformer-based models for feature extraction and semantic embedding generation. The overall pipeline consists of three main stages: (i) modality-specific feature extraction, (ii) multimodal embedding construction, and (iii) cross-modal similarity computation.

3.4.1 *Music Representation Learning.* For music, we extract two distinct feature representations. First, we learn acoustic features using MERT similar to the contrastive approach in Mozart (Section 3.2) to obtain deep audio embeddings, denoted as $\mathbf{m}_{\text{MERT}} \in \mathbb{R}^{d_m}$. Next, we learn textual representations of music's semantic description by querying an LLM (GPT-4o) with metadata including the V-A values. The resulting textual descriptions were then converted to an embedding using the transformer-based language model BERT, yielding $\mathbf{m}_{\text{LLM}} \in \mathbb{R}^{d_t}$.

The final music embedding is obtained by concatenating these two representations:

$$\mathbf{m} = [\mathbf{m}_{\text{MERT}}; \mathbf{m}_{\text{LLM}}] \in \mathbb{R}^{d_m + d_t} \tag{9}$$

To ensure a compact representation, we train an autoencoder $f_m : \mathbb{R}^{d_m + d_t} \rightarrow \mathbb{R}^{256}$ that reduces the dimensionality of $\mathbf{m}$ to $\mathbf{z}_m \in \mathbb{R}^{256}$.

3.4.2 *Painting Representation Learning.* For paintings, a similar two-stream feature extraction strategy is employed. Similar to Mozart, we extracted visual features of paintings using ResNet-50, obtaining $\mathbf{p}_{\text{ResNet}} \in \mathbb{R}^{d_p}$. Subsequently for textual representation, we queried a VLM model (GPT-4o) with the painting images, their metadata, valence-arousal values, and emotion labels from WikiArt. The textual descriptions from GPT-4o were then converted to an embedding using the transformer-based language model BERT to produce $\mathbf{p}_{\text{VLM}} \in \mathbb{R}^{d_v}$.

The overall painting embedding is constructed by concatenation:

$$\mathbf{p} = [\mathbf{p}_{\text{ResNet}}; \mathbf{p}_{\text{VLM}}] \in \mathbb{R}^{d_p + d_v} \tag{10}$$

As with music, we employ an autoencoder $f_p : \mathbb{R}^{d_p + d_v} \rightarrow \mathbb{R}^{256}$ to obtain a reduced representation $\mathbf{z}_p \in \mathbb{R}^{256}$.

3.4.3 *Cross-Modal Similarity Computation.* Given the 256D embeddings $\mathbf{z}_m$ and $\mathbf{z}_p$ for music and paintings, respectively, we compute the similarity matrix $S_{LV} \in \mathbb{R}^{N_m \times N_p}$ between all music and painting pairs using cosine similarity. As in Mozart, a weighted average is computed to make recommendations based on user preference ratings using this similarity matrix:

$$d_{\text{Salieri}}(p_j) = \sum_{i=1}^{k} w_i \cdot \|\mathbf{z}_{m_i} - \mathbf{z}_{p_j}\|_2 \tag{11}$$

This approach allows us to compare LLM and VLM-based representations with both the affective space search and contrastive alignment techniques in CDR for AT.

Algorithm 1: CDR preprocessing pipeline

1: **procedure** PreprocessCDR(M, P)
2:   $x_m \leftarrow$ ExtractMERT(M)   // Mozart music features
3:   $x_p \leftarrow$ ExtractResNet50(P)   // Mozart VA features
4:   $\mathbf{v}_m, \mathbf{v}_p \leftarrow$ GetVAScores(M, P)   // Haydn
5:   $z_m^* \leftarrow$ Reduce $(x_m, 256\text{D}) + \mathbf{v}_m \rightarrow \mathbb{R}^{258}$ // Mozart
6:   $z_p^* \leftarrow$ Reduce $(x_p, 256\text{D}) + \mathbf{v}_p \rightarrow \mathbb{R}^{258}$ // Mozart
7:   $z_m^+, z_p^+ \leftarrow$ Reduce $(z_m^*, z_p^*, 128\text{D})$ // Mozart
8:   $m_{\text{MERT}} \leftarrow$ ExtractMERT(M)   // Salieri music features
9:   $m_{\text{LM}} \leftarrow$ ExtractLLM(M)   // Salieri music text features
10:   $p_{\text{ResNet}} \leftarrow$ ExtractResNet50(P)   // Salieri VA features
11:   $p_{\text{VLM}} \leftarrow$ ExtractVLM(P)   // Salieri VA text features
12:   $m \leftarrow [m_{\text{MERT}}; m_{\text{LM}}]$ // Salieri concatenation
13:   $p \leftarrow [p_{\text{ResNet}}; p_{\text{VLM}}]$ // Salieri concatenation
14:   $z_m, z_p \leftarrow$ Reduce $(m, p, 256\text{D})$ // Salieri
15:   **return** $\mathbf{v}_m, \mathbf{v}_p, z_m^+, z_p^+, z_m, z_p, p_{\text{ResNet}}$

Algorithm 2: CDR recommendations

Precondition: Extracted features for each engine (Algorithm 1).
1: **procedure** RecommendCDR(M, P, R)
2:   $L \leftarrow d(p_j)$ using ratings $w_i \in R$ // See (7) (8) (11)
3:   Sort(L)   // Ascending order
4:   **return** Slice(L, n)   // Top-n paintings

| (a) Pre-study questions | (b) Preference elicitation | (c) Guided AT session | (d) Post-study questions |

Figure 2: Screenshots (in chronological order) of the web application we developed for the user study.

4 User Study

We conducted a large-scale online study to understand the user's perception towards the quality of our CDR strategies for psychiatric sequelae rehabilitation therapy, and ultimately to assess their efficacy in supporting their healing journey. Hence, we designed a guided AT study that follows the setup proposed in SOTA VA RecSys for digital AT [49, 50]. In addition to the three CDR engines, we developed a SOTA Visual engine that produces recommendations based on elicited painting preferences using cosine similarity over ResNet50-extracted features.

4.1 Materials

We curated the music and painting samples from DEAM and WikiArt to ensure that they were suitable for conducting guided AT sessions. Music and paintings were retained if their valence score exceeded 0.1, ensuring positive emotionality as per prior therapeutic studies [7, 40], and arousal was either ≤ -0.1 or ≥ 0.1, excluding neutral samples to enhance emotional engagement [41, 48].

The curated music set supported preference elicitation, while paintings were matched with a set of nature-based artworks that were previously validated to support AT [49]. This allowed us to identify paintings in WikiArt closely matching those used in digital AT. Further, to ensure a safe deployment, we conducted a pilot test with a registered art therapist[7] to manually go through the curated paintings and remove potentially harmful or unsuitable paintings, yielding a final 239×63 similarity matrix (239 music tracks, 63 paintings).

4.2 Participants

We recruited 200 people who had psychiatric sequelae via the Prolific crowdsourcing platform.[8] All participants indicated that they were officially diagnosed with any of the following mental health conditions: Anxiety (165 participants), Depression (162), Attention Deficit Hyperactivity Disorder (70), Bipolar Disorder (45), Eating Disorder (36), Post-traumatic Stress Disorder (28), among other conditions. We also administered the Patient Health Questionnaire-4 (PHQ-4) [31] to collect explicit signs of anxiety and depression. All participants showed symptoms for both of these conditions. Participants also indicated to have been through any of the following treatments: Psychological therapy (188), Medication (137), Self-care (71), Hospital and residential treatment programs (47), and Brain-stimulation treatments (27). Furthermore, we enforced other criteria for eligibility: being fluent in English, minimum approval rate of

[7]https://www.arttherapyfederation.eu/
[8]https://www.prolific.com/

100% in previous crowdsourcing studies in Prolific, and being active in the last 90 days.

Our recruited participants (100 female, 100 male) were aged 31.4 years (SD=10.7) and could complete the study only once. Most of them lived in South Africa (85), Poland (21), United Kingdom (16), and Canada (12). Participants had been through planned or emergency surgical procedures (88), traumatic experience such as injuries from an accident or a fall (68), or medical conditions such as pneumonia, heart failure, or stroke (57). Participants had been in a general ward or medical/surgical unit (85), ICU (73), or in an Emergency Room (54). For most participants, the duration of their stay in a hospital was less than a week (89), between 1 to 2 weeks (47), or between 2 to 4 weeks (45). The study took a mean time of 35 min (SD=15) to complete. Participants were paid an equivalent hourly wage of $12/h.

4.3 Apparatus

We designed a web application (Figure 2) that first assessed baseline affective states, then elicited music and painting preferences,[9] then facilitated a guided AT session, and finally assessed post-test affective states. We indicated in Prolific that participants had to use a desktop-based computer, to ensure they would be in a comfortable environment, and that the study would require them to use headphones.

4.4 Design

We deployed the three CDR engines (Mozart, Haydn, and Salieri) together with the SOTA baseline VA RecSys engine (Visual). Each participant was only exposed to one of the four engines (between-subjects design) and then went through a session of guided AT comprising three paintings (the top-3 recommendations from each engine, based on the elicited user preferences, see 2). Each group comprised a gender-balanced set of 50 participants.

4.5 Procedure

We elicited user preferences by providing a set of 11 music tracks and 11 paintings, randomly sampled from DEAM and WikiArt. Participants had to rate each item in a 1–5 scale. One music track and one painting were attentional checks, which had to be rated as 1 and were not used by any engine. Participants had to listen each music track for 15 seconds.

We assessed pre-test and post-test affective states using the Pick-A-Mood tool [10] and the short version of the Positive and Negative Affect Schedule (PANAS) scale [47], which considers 5 positive, 5

[9]Paintings were necessary for the Visual engine only.
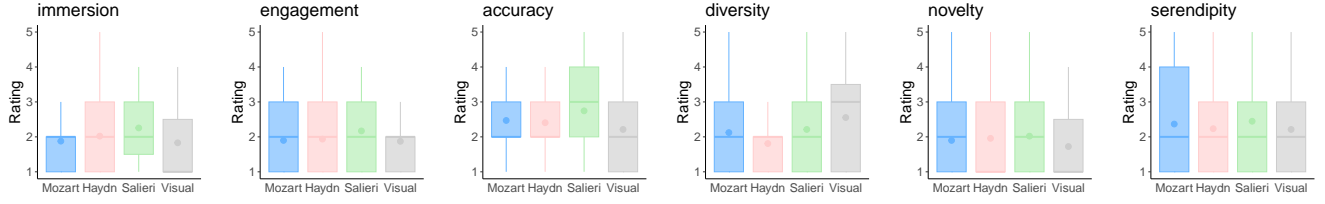
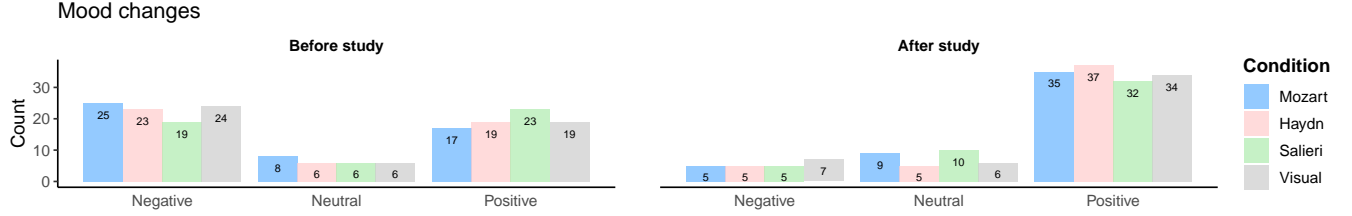**Figure 3: Distribution of ratings of our user-centric recommendation quality metrics. Dots denote mean values.**

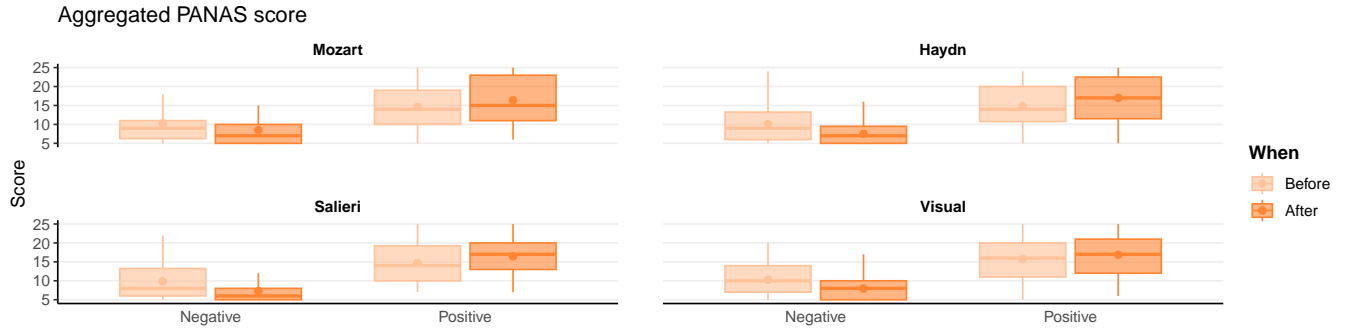**Figure 4: Mood improvement comparison before and after art therapy.**

Figure 5: Emotion score changes according to the Positive Affect Negative Affect Schedule (PANAS) scale.

negative, and 1 neutral item to assess the affective state of participants before and after guided AT.

The guided AT session comprised three paintings that were recommended to each participant based on the RecSys engine they were assigned to. Participants were encouraged to engage with the paintings: *"Imagine yourself entering the painting and exploring it. How did you feel while spending time in this painting?"*. They were prompted to reflect on their healing experience and describe it in a few sentences. They were also prompted to describe which aspects of the painting contributed most to their experience.

After the AT session, in addition to the affective states self-assessment, participants were asked to rate the recommendations in a 5-point Likert scale according to widely accepted proxies of recommendation quality [37, 49]:

Accuracy: To what degree did the recommended paintings match your selected personal preference?
Diversity: To what extent did you find the recommended paintings diverse?
Novelty: To what extent did the recommendations help you discover paintings you didn't know before?

Serendipity: To what extent did you find surprisingly interesting paintings among the recommendations?
Immersion: How much did the recommended paintings contribute to your sense of immersion, making you feel deeply involved or absorbed in the artwork?
Engagement: To what extent did the recommended paintings contribute to your feeling of engagement, capturing your attention and generating a sense of involvement or interest?

4.6 Results

Our analysis is structured in terms of user-centric metrics of recommendation quality, emotional changes post-therapy, and user reflections.

4.6.1 Analysis of recommendation quality measures. We used a linear mixed-effects (LME) model where participants are considered random effects.[10] We fitted one LME model per user-centric metric and computed the estimated marginal means for specified factors. We then ran pairwise comparisons (also known as *contrasts*) with Bonferroni-Holm correction to guard against multiple comparisons.

---

[10]An LME model is appropriate because ratings are discrete and have a natural order.

Figure 3 shows the distributions of user ratings for the user-centric metrics of recommendation quality. We did not find statistically significant differences between engines, suggesting that all engines were rated similarly by our participants. The only exception is Visual outperforming Haydn in terms of diversity ($p = .007$) with a moderate effect size ($r = 0.2$).

4.6.2 Analysis of changes in mood. Figure 4 shows the mood changes before and after therapy, for the four recommendation groups we have considered in our study. In all groups, a mood enhancement effect of guided art therapy was observed. When comparing it to the baseline, where many participants were in a negative mood (46.6%), after guided art therapy the majority of participants reported being in a positive mood (72.6%), with only a minority remaining in negative (11.5%) or neutral (15.7%) mood. The differences between approaches were not statistically significant: $\chi^2(3, N = 195) = 1.08, p = .780, \phi = 0.075$.

Figure 5 shows the change in scores after therapy, aggregated according to the ten items of the PANAS scale. Differences between groups were not statistically significant in any case, with small to moderate effect sizes.

4.6.3 Analysis of user reflections. We conducted a thematic analysis to identify the most salient healing themes from the user reflections, which were submitted as free-form comments during the AT session. Starting from a list of 10 themes (with sample quotes from participants) that follow established therapeutic practice [50], we seeded BERTopic [16] with keywords derived from each theme's description and sample quote. For example, for the Safety theme, the following keywords were identified: "safe", "secure", "protected", "calm", "familiar". Then, each user reflection towards a particular painting was assigned a dominant theme based on topic coherence, with mappings refined by examining the top words per topic. The resulting theme frequencies are reported in Figure 6.
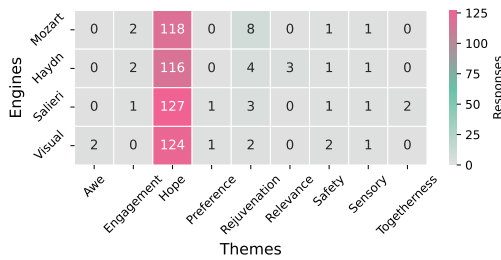

Figure 6: Heatmap of healing concepts identified by recommendation engines using guided topic modeling.

Notably, all engines predominantly elicited Hope, with Salieri showing the highest focus (127 responses), followed by Visual (124), Mozart (118), and Haydn (116). Again, compared to the SOTA image-based engine, the music-based approaches demonstrate a competitive therapeutic profile, even outperforming SOTA in most user-centric metrics of recommendation quality, suggesting that music-based preference elicitation methods are an effective alternative in personalized AT.

5 Discussion

We have made the first step in digital AT personalization by introducing affect-aware music-driven preference elicitation. Taken together, our CDR algorithms (Mozart, Haydn, and Salieri)[11] unlock new possibilities, mapping music's affective cues to therapeutic VA recommendations. The findings from our study reveal distinct strengths of our approaches. For example, Salieri excels in promoting Hope (127 responses, Figure 6) while Haydn delivers interpretable affective alignment. This suggests that music-driven elicitation offers focused emotional resonance. Furthermore, mood improvements (46.6% negative to 72.6% positive post-therapy, Figure 4) across all proposed CDR algorithms also signify music's broader emotional appeal. All in all, these findings suggest CDR holds potential to open new avenues for digital AT via multimodal frameworks. For example, AT systems could connect to music streaming applications like Spotify or Dizzer, thus leveraging the users' listening history or playlists to derive therapeutic recommendations.

6 Limitations and Future Work

A primary limitation in this work is the scarcity of therapeutically curated painting collections with reliable affective labels. Prior work [49, 50] highlighted the potential harm of unfiltered visual stimuli in digital AT, necessitating therapist intervention to safeguard recommendations. We align with this view, yet the lack of parallel datasets, affectively labeled and within therapeutic valence-arousal ranges, poses a significant challenge to scaling music-driven VA recommendations. Future efforts could address this through crowdsourced studies, similar to WikiArt Emotions, to curate robust therapeutic datasets, fostering a research space where affect-aware CDR algorithms can excel. Furthermore, the lack of significant differences between systems may be attributed to stimulus homogeneity, as a result of filtered samples implemented as safety measures to prevent the exposure of patients to negative stimuli. We believe longitudinal studies to further validate the therapeutic impacts of these approaches are exciting avenues for future work.

7 Conclusion

We have introduced affect-aware cross-domain recommendation for digital art therapy by leveraging music preference elicitation to personalize visual art recommendations. Our family of methods (Mozart, Haydn, Salieri) bridge the music-art gap, enhancing emotional engagement through cross-modal affective transfer. Our study showed these approaches match state-of-the-art visual baselines in metrics like accuracy and diversity while significantly improving mood and positive affect. This work advances RecSys by addressing single-domain limitations, and we hope to inspire multimodal RecSys for well-being.

Acknowledgments

---

[11]Mozart, Haydn, and Salieri represent not just classical music composers but contrasting philosophies (brilliance, balance, and rivalry). These names were chosen to mirror the unique design principles behind each recommendation method.

References

[1] Annemarie Abbing, Anne Ponstein, Susan van Hooren, Leo de Sonneville, Hanna Swaab, and Erik Baars. 2018. The effectiveness of art therapy for anxiety in adults: A systematic review of randomised and non-randomised controlled trials. *PloS one* 13, 12 (2018), e0208716.

[2] Sinem Aslan, Giovanna Castellano, Vincenzo Digeno, Giuseppe Migailo, Raffaele Scaringi, and Gennaro Vessio. 2022. Recognizing the emotions evoked by artworks through visual features and knowledge graph-embeddings. In *International Conference on Image Analysis and Processing*. Springer, 129–140.

[3] Kelly Sarah Barnett and Fabian Vasiu. 2024. How the arts heal: a review of the neural mechanisms behind the therapeutic effects of creative arts on mental and physical health. *Frontiers in behavioral neuroscience* 18 (2024), 1422361.

[4] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[5] Rebecca Bokoch, Noah Hass-Cohen, April Espinoza, Tyler O'Reilly, and Elad Levi. 2025. A scoping review of integrated arts therapies and neuroscience research. *Frontiers in Psychology* 16 (2025), 1569609.

[6] Olivia Brancatisano, Amee Baird, and William Forde Thompson. 2020. Why is music therapeutic for neurological disorders? The Therapeutic Music Capacities Model. *Neuroscience & Biobehavioral Reviews* 112 (2020), 600–615.

[7] José J Campos-Bueno, Octavio DeJuan-Ayala, Pedro Montoya, and Niels Birbaumer. 2015. Emotional dimensions of music and painting and their interaction. *The Spanish journal of psychology* 18 (2015), E54.

[8] Suvin Choi, Jong-Ik Park, Cheol-Ho Hong, Sang-Gue Park, and Sang-Cheol Park. 2024. Accelerated construction of stress relief music datasets using CNN and the Mel-scaled spectrogram. *PloS one* 19, 5 (2024), e0300607.

[9] Shreyan Chowdhury, Andreu Vall, Verena Haunschmid, and Gerhard Widmer. 2019. Towards Explainable Music Emotion Recognition: The Route via Mid-level Features. In *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019, Delft, The Netherlands, November 4-8, 2019*, Arthur Flexer, Geoffroy Peeters, Julián Urbano, and Anja Volk (Eds.). 237–243. http://archives.ismir.net/ismir2019/paper/000027.pdf

[10] Pieter MA Desmet, Martijn H Vastenburg, and Natalia Romero. 2016. Mood measurement with Pick-A-Mood: review of current methods and design of a pictorial self-report scale. *Journal of Design Research* 14, 3 (2016), 241–279.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[12] Antonella Di Vita, Mario Augusto Procacci, Martina Bellagamba, Maria Jacomini, Roberta Massicci, and Maria Paola Ciurli. 2022. Psychotherapy and Art Therapy: A pilot study of group treatment for patients with traumatic brain injury. *Journal of health psychology* 27, 4 (2022), 836–846.

[13] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th international conference on world wide web*. 278–288.

[14] Yingjie Feng and Mingda Wang. 2025. Effect of music therapy on emotional resilience, well-being, and employability: a quantitative investigation of mediation and moderation. *BMC psychology* 13, 1 (2025), 47.

[15] Ignacio Fernández-Tobías, Matthias Braunhofer, Mehdi Elahi, Francesco Ricci, and Iván Cantador. 2016. Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction* 26 (2016), 221–255.

[16] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

[17] Suzanne Haeyen and Merel Staal. 2021. Imagery rehearsal based art therapy: Treatment of post-traumatic nightmares in art therapy. *Frontiers in psychology* 11 (2021), 628717.

[18] Kathy Hathorn and Upali Nanda. 2008. A guide to evidence-based art. *The Center for Health Design* 1 (2008), 1–23.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[20] Taisei Hirakawa, Keisuke Maeda, Takahiro Ogawa, Satoshi Asamizu, and Miki Haseyama. 2021. Cross-domain recommendation method based on multi-layer graph analysis with visual information. In *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2688–2692.

[21] Jingxuan Hu, Jinhuan Zhang, Liyu Hu, Haibo Yu, and Jinping Xu. 2021. Art therapy: a complementary treatment for mental disorders. *Frontiers in psychology* 12 (2021), 686005.

[22] Papadopoulos Stefanos Iordanis. 2021. Emotion-aware music recommendation systems. (2021).

[23] Dian Jiao. 2025. Advancing personalized digital therapeutics: integrating music therapy, brainwave entrainment methods, and AI-driven biofeedback. *Frontiers in Digital Health* 7 (2025), 1552396.

[24] Patrik N. Juslin and Daniel Västfjäll. 2008. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences* 31, 5 (2008), 559–575. https://doi.org/10.1017/S0140525X08005293

[25] Stefan Koelsch. 2014. Brain correlates of music-evoked emotions. *Nature reviews neuroscience* 15, 3 (2014), 170–180.

[26] AI Ladas, T Gravalas, C Katsoridou, and CA Frantzidis. 2024. Harmony in the brain: A narrative review on the shared neural substrates of emotion regulation and creativity. *Brain Organoid and Systems Neuroscience Journal* 2 (2024), 81–91.

[27] Cheng-Che Lee, Wan-Yi Lin, Yen-Ting Shih, Pei-Yi Kuo, and Li Su. 2020. Crossing you in style: Cross-modal style transfer from music to visual arts. In *Proceedings of the 28th ACM international conference on multimedia*. 3219–3227.

[28] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. *arXiv preprint arXiv:2201.12086* (2022).

[29] Yizhi Li, Ruibin Yuan, Ge Zhang, Yinghao Ma, Xingran Chen, Hanzhi Yin, Chenghao Xiao, Chenghua Lin, Anton Ragni, Emmanouil Benetos, et al. 2023. Mert: Acoustic music understanding model with large-scale self-supervised training. *arXiv preprint arXiv:2306.00107* (2023).

[30] Bojun Liu and Bohong Liu. 2022. Human-Centric C ross-Domain T ransfer N etwork for M usic Recommendation. In *International Advanced Computing Conference*. Springer, 407–414.

[31] Bernd Löwe, Inka Wahl, Matthias Rose, Carsten Spitzer, Heide Glaesmer, Katja Wingenfeld, Antonius Schneider, and Elmar Brähler. 2010. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of affective disorders* 122, 1-2 (2010), 86–95.

[32] Cathy A Malchiodi. 2011. *Handbook of art therapy*. Guilford Press.

[33] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-domain recommendation: An embedding and mapping approach. In *IJCAI*, Vol. 17. 2464–2470.

[34] Saif M Mohammad. 2025. NRC VAD Lexicon v2: Norms for valence, arousal, and dominance for over 55k English terms. *arXiv preprint arXiv:2503.23547* (2025).

[35] Sharaj Panwar, Paul Rad, Kim-Kwang Raymond Choo, and Mehdi Roopaei. 2019. Are you emotional or depressed? Learning about your emotional state from your music using machine learning. *The Journal of Supercomputing* 75 (2019), 2986–3009.

[36] Varesh Patel and Khyati Mehta. 2024. Emotion-Aware Music Recommendations: Evaluating Custom CNN vs. VGG16 and InceptionV3. In *2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS)*. IEEE, 499–504.

[37] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. 157–164.

[38] Alfredo Raglio, Marcello Imbriani, Chiara Imbriani, Paola Baiardi, Sara Manzoni, Marta Gianotti, Mauro Castelli, Leonardo Vanneschi, Francisco Vico, and Luca Manzoni. 2020. Machine learning techniques to predict the effectiveness of music therapy: A randomized controlled trial. *Computer Methods and Programs in Biomedicine* 185 (2020), 105160. https://doi.org/10.1016/j.cmpb.2019.105160

[39] Yiren Ren, Sophia Kaltsouni Mehdizadeh, Grace Leslie, and Thackery Brown. 2024. Affective music during episodic memory recollection modulates subsequent false emotional memory traces: an fMRI study. *Cognitive, Affective, & Behavioral Neuroscience* 24, 5 (2024), 912–930.

[40] Mathieu Roy, Isabelle Peretz, and Pierre Rainville. 2008. Emotional valence contributes to music-induced analgesia. *Pain* 134, 1-2 (2008), 140–147.

[41] Gillian M Sandstrom and Frank A Russo. 2010. Music hath charms: The effects of valence and arousal on recovery following an acute stressor. *Music and Medicine* 2, 3 (2010), 137–143.

[42] Susanne Schweizer, Ian H Gotlib, and Sarah-Jayne Blakemore. 2020. The role of affective control in emotion regulation during adolescence. *Emotion* 20, 1 (2020), 80.

[43] Heather L Stuckey and Jeremy Nobel. 2010. The connection between art, healing, and public health: A review of current literature. *American journal of public health* 100, 2 (2010), 254–263.

[44] J. Tan et al. 2024. Contrastive Learning Is Spectral Clustering On Similarity Graph. In *Proceedings of CVPR 2024*.

[45] Michael Thaut. 2013. *Rhythm, music, and the brain: Scientific foundations and clinical applications*. Routledge.

[46] Berit Marie Dykesteen Vik, Geir Olve Skeie, and Karsten Specht. 2019. Neuroplastic effects in patients with traumatic brain injury after music-supported therapy. *Frontiers in human neuroscience* 13 (2019), 177.

[47] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

[48] Barbara L Wheeler, Estate Sokhadze, Joshua Baruth, Gene Ann Behrens, and Carla F Quinn. 2011. Musically induced emotions: Subjective measures of arousal and valence. *Music and Medicine* (2011).

[49] Bereket A. Yilma, Chan Mi Kim, Gerald C. Cupchik, and Luis A. Leiva. 2024. Artful Path to Healing: Using Machine Learning for Visual Art Recommendation to Prevent and Reduce Post-Intensive Care Syndrome (PICS). In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 447, 19 pages. https://doi.org/10.1145/3613904.3642636

[50] Bereket A Yilma, Chan Mi Kim, Geke Ludden, Thomas van Rompay, and Luis A Leiva. 2025. The AI-Therapist Duo: Exploring the Potential of Human-AI Collaboration in Personalized Art Therapy for PICS Intervention. *arXiv preprint arXiv:2502.09757* (2025).

[51] Saba Yousefian Jazi, Marjan Kaedi, and Afsaneh Fatemi. 2021. An emotion-aware music recommender system: bridging the user's interaction and music recommendation. *Multimedia Tools and Applications* 80 (2021), 13559–13574.

[52] Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. 2022. A survey on cross-domain recommendation: taxonomies, methods, and future directions. *ACM Transactions on Information Systems* 41, 2 (2022), 1–39.

[53] Qian Zhang, Wenhui Liao, Guangquan Zhang, Bo Yuan, and Jie Lu. 2021. A deep dual adversarial network for cross-domain recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 4 (2021), 3266–3278.

[54] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-domain recommendation: challenges, progress, and prospects. *arXiv preprint arXiv:2103.01696* (2021).